INTRODUCTION TO MATRIX GENERALIZED INVERSES AND THEIR APPLICATIONS

By

SHELDON E. NEWHOUSE
NASA, Goddard Space Flight Center
Greenbelt, Maryland

PREFACE

This paper is an attempt at an elementary presentation of what the author considers to be the most applicable and interesting properties of generalized inverses of matrices. We shall pay particular attention to a discussion of those properties concerned with the pseudo-inverse of a rectangular matrix. This structure was first studied by Moore, who used the name general reciprocal, and later it was studied by Penrose, who used the name generalized inverse.

The present discussion is what I deem to be the most practical way to present this material with the expectation that some of the readers of this paper will not be very well acquainted with abstract mathematics. I should like to apologize if all credit for work is not properly placed and refer those interested readers to the list of references, where practically all the material presented here is discussed in detail.

I wish to thank all the people who were involved in the preparation and presentation of this paper. In particular, I should like to express my gratitude to E. R. Lancaster, who proof-read the manuscript and made many valuable suggestions, and to C. A. Rohde, who was kind enough to let me read his doctoral thesis (Ref. 25), and whose suggestions were of immeasurable aid to me in the compilation and presentation of this material.

I. PRELIMINARIES

To begin, we shall recall some results and definitions from matrix theory. If n and m are natural numbers, an n x m matrix A will be a rectangular array of real numbers

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ & & \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}$$

which will be denoted by $[a_{ij}]$, or A_{ij}, $1 \le i \le n$, $1 \le j \le m$. The transpose of A, denoted by A', will mean the m x n matrix whose (i,j)th entry is a_{ji}. The i<u>th</u> row of A is the 1 x m matrix $[a_{i1} \; . \; . \; . \; a_{im}]$. The j<u>th</u> column of A is the n x 1 matrix

$$\begin{bmatrix} a_{1j} \\ \cdot \\ \cdot \\ \cdot \\ a_{nj} \end{bmatrix}$$

Thus, the transpose A' of A is the matrix whose columns are the rows of A. We note that for matrices A and B whose product is defined, (AB)' = B'A', and for A and B whose sum is defined, (A+B)' = A'+B'. An m-row vector will mean a 1 x m matrix, and an m-column vector will mean an m x 1 matrix. When the meaning is clear from the context, an m-row or column vector will simply be called an m-vector, or vector. Note that if

x is an m-column vector, then x' is an m-row vector and vice versa. In what follows, unless otherwise stated, we shall always mean that x is a column vector, and we shall write x' when we wish to refer to the corresponding row vector.

Let A, B be $n \times m$ matrices with $A = [a_{ij}]$ and $B = [b_{ij}]$. We define the $n \times m$ matrix $A+B$ by $(A+B)_{ij} = [a_{ij} + b_{ij}]$. If A is $n \times m$, B is $m \times p$, we define the $n \times p$ matrix $C = AB$ by

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}$$

The $n \times m$ matrix $\bar{0}$ is the matrix all of whose entries are zeros. The $n \times n$ matrix I_n is the n-identity matrix $[\delta_{ij}]$ where

$$\delta_{ij} = \begin{matrix} 0 \text{ if } i \neq j \\ 1 \text{ if } i = j \end{matrix}$$

Thus $$I_n = \begin{bmatrix} 1 & 0 & . & . & . & 0 \\ 0 & 1 & . & . & . & 0 \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & 1 \end{bmatrix}$$

The $n \times n$ matrix D is called a diagonal matrix, denoted by $D = \text{diag}\ \{a_i, \ldots, a_n\}$ if the non-diagonal entries are zeros, and $D_{ii} = a_i$, $1 \leq i \leq n$.

$$\text{Thus } D = \text{diag}\ \{a_i, \ldots, a_n\} = \begin{bmatrix} a_1 & 0 & .. & 0 \\ 0 & a_2 & & . \\ . & . & & . \\ . & & . & . \\ . & & . & . \\ 0 & . & & a_n \end{bmatrix}$$

The $n \times n$ matrix A is called symmetric if $A = A'$, and idempotent if $A^2 = A$.

Let A be an n x m matrix. An n x n matrix B such that BA = A is called a left identity for A, while an m x m matrix T such that AT = A is called a right identity for A. The n x n matrix A is called invertible or non-singular if there is an n x n matrix B such that $AB = BA = I_n$. When such a matrix B exists, it is unique and denoted by A^{-1}. A^{-1} is called the inverse of A. If the matrix A is not invertible, it is called singular. A matrix P such that $P^{-1} = P'$ is called an orthogonal, or orthonormal matrix.

We shall need the following result which we state without proof (See Ref. 3).

Theorem I.1: Let A be an n x n symmetric matrix. Then there is an orthogonal matrix P such that

$$PAP' = D = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_r, 0, \ldots, 0\}$$

where the λ_i are non-zero scalars which may or may not be distinct.

The matrix R with s non-zero rows is said to be in row reduced echelon form if the following are satisfied:

i) If a row is not all zero, its leading non-zero term is 1.

ii) All of the non-zero rows are above the rows consisting only of zeros.

iii) If the $i\underline{th}$ non-zero row has its leading non-zero term in the $j\underline{th}$ column, then the leading non-zero term in each row above row i appears in a column to the left of column j; for $1 \le i \le s$

iv) If the leading non-zero term in the $i\underline{th}$ row appears in the $j\underline{th}$ column, then the other entries of the $j\underline{th}$ column are all zeroes.

The matrix C is said to be in column reduced echelon form if C' is in row reduced echelon form. We may sometimes speak of R as being row reduced or in row reduced form.

It is well-known (see Ref. 17) that any n x m matrix A can be put into a unique row reduced form by elementary row operations. If R is this row-reduced form of A, there is a non-singular n x n matrix P such that PA = R.

Similarly, we may use elementary column operations to obtain a column reduced form B of R. When this is done, it is seen that there is a non-singular m x m matrix Q such that

$$PAQ = B = \begin{bmatrix} I_r & \bar{0} \\ \bar{0} & \bar{0} \end{bmatrix}$$

where the $\bar{0}$'s stand for zero matrices of appropriate sizes and $r \le \min\{m,n\}$.

We shall occasionally have use for so-called partitioned matrices. These are matrices whose entries are themselves matrices. We observe that, assuming all operations are defined, these matrices may be handled as though the entries were scalars. When using partitioned matrices, we may use lines to separate the submatrix entries. For example, if $P = \left[\begin{array}{c} P_1 \\ \hline P_2 \end{array}\right]$ where P_1 is an $n_1 \times n_2$ matrix, P_2 is an $n_3 \times n_2$ matrix and $Q = [Q_1 \mid Q_2]$ where Q_1 is an $n_2 \times n_4$ matrix, Q_2 is an $n_2 \times n_5$ matrix, then we have PQ is an $(n_1+n_3) \times (n_4+n_5)$ matrix given by

$$PQ = \left[\begin{array}{c} P_1 \\ \hline P_2 \end{array}\right] [Q_1 \mid Q_2] = \begin{bmatrix} P_1Q_1 & P_1Q_1 \\ P_2Q_1 & P_2Q_2 \end{bmatrix}$$

If A is an $n \times n$ matrix, the trace of A is defined by

$$\mathrm{tr}\, A = \sum_{i=1}^{n} A_{ii}$$

The trace has the following properties:

i) For matrices A, B such that AB and BA exist, $\mathrm{tr}(AB) = \mathrm{tr}(BA)$. Hence, $\mathrm{tr}(PAP^{-1}) = \mathrm{tr}\, A$, where P is non-singular and PAP^{-1} exists.

ii) $\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$ where $A+B$ exists.

A finite set of vectors $\{x_i, \ldots, x_n\}$ is called linearly dependent if there are n real numbers $c_1, \ldots, c_n$, not all zero, such that $\sum_{i=1}^{n} c_i x_i = 0$.

If $\{x_i\}_1^n$ are not linearly dependent, they are said to be linearly independent.

A row or column vector space V is a collection of row or column vectors such that

i) $x, y \in V \Rightarrow x+y \in V$ and

ii) $x \in V$ and c, a real number $\Rightarrow cx \in V$.

A subspace W of a vector space V is a subset of V which is itself a vector space. Thus, for example, the collection of all n-column vectors (n is fixed) whose first coordinate is zero is a subspace of the vector space consisting of all n-column vectors.

If $\{x_i\}_1^n$ is a finite collection of vectors, a vector x is said to be a linear combination of the vectors $x_1, \ldots, x_n$ if there are n real numbers $c_1, \ldots, c_n$ such that $x = \sum_{i=1}^{n} c_i x_i$. The collection V of all vectors which are linear combinations of $x_1, \ldots, x_n$ forms a vector space, $V = L\{x_1, \ldots, x_n\}$, called the space spanned by the vectors $x_1, \ldots, x_n$. A linearly independent spanning set of vectors for a vector space V is called a basis for V. It can be proved (Ref. 17) that every vector space has a basis and that any two bases for the same vector space have the number of elements. In all of our discussions involving a vector space V, we shall assume that there is a finite basis for V. We shall say that V is finite dimensional and that the dimension of V, dim V, is n, where n is the number of elements in a basis for V.

For example, let V be the set of all n-column vectors. It is clear that the n-vectors $e_j = [0, \ldots, 0, 1_j, 0, \ldots]'$ form a basis for V, and that this basis has exactly n elements.

Now we return to an n x m matrix A. The row (column) space of A is the vector space spanned by the rows (columns) of A. The dimension of the row (column) space of A is called the row (column) rank of A. It can be shown that for any matrix, row rank A = column rank A. This common number is called the rank of A and denoted by rk A. This shows that for any matrix A, rk A = rk A'.

If x and y are n-column vectors, the inner or dot product, (x,y), of x and y is the number α so that $x'y = [\alpha]$. In this situation we may choose to identify the matrix $[\alpha]$ with the real number α itself and write $x'y = \alpha$.

The inner product has the following properties:

i) $(x, y) = (y, x)$

ii) $(x, y) = (x, \alpha y) = (\alpha x, y) = \alpha(x, y)$

iii) $(x_1 + x_2. \ y) = (x_1, y) + (x_2, y)$

iv) $(x, y_1 + y_2) = (x, y_1) + (x, y_2)$

v) $(x, x) \geq 0$, and $(x, x) = 0 \Rightarrow x = \bar{0}$.

Here x, x_1, x_2, y, y_1, y_2 are all vectors of the same dimension, and α is a real number. Property i) is called commutativity. Properties ii) - iv) characterize the inner product as being bilinear function. Property v) is called positive definiteness. For a vector x, the norm or magnitude of x is the number $(x, x)^{\frac{1}{2}}$. This is denoted by $\|x\|$ and has the following properties:

i) $\| x \| \geq 0$ and $\| x \| = 0 \Rightarrow x = \bar{0}$

ii) $\| \alpha x \| = | \alpha | \, \| x \|$

iii) $\| x + y \| \leq \| x \| + \| y \|$

Let A be an $n \times m$ matrix, y be a fixed n-vector. Consider the numbers $\|Ax-y\|$ as x varies through all m-vectors. We define the unique number $\inf_x \|Ax-y\|$ to be the number α such that

i) $\alpha \leq \|Ax-y\|$ for all m-vectors x.

ii) If β is any number so that $\beta \leq \|Ax-y\|$ for all m-vectors x, then $\beta \leq \alpha$.

Two vectors x and y are called orthogonal if $(x, y) = 0$. We write $x \perp y$. A vector x is orthogonal to a vector space V if for all $y \in V$, $x \perp y$. We write $x \perp V$. Similarly, two vectors spaces V and W are orthogonal if, for all $x \in V$, $y \in W$, $x \perp y$.

We will need the following well-known results:

Theorem I.2. If A is an $n \times m$ matrix, B an $m \times p$ matrix, then $\text{rk } AB \leq \text{rk } A$, and $\text{rk } AB \leq \text{rk } B$.

Proof: Let us first note that for any matrices A and B whose product is defined, the rows of AB are in the row space of B, and the columns of AB are in the column space of A.

Now, if W and V are vector spaces with $W \subset V$, then $\dim W \leq \dim V$. Hence,

$\text{rk } AB = \text{row rk } AB \leq \text{row rk } B = \text{rk } B$ and

$\text{rk } AB = \text{col rk } AB \leq \text{col rk } A = \text{rk } A$. Q.E.D.

If A is an $n \times m$ matrix, the range of A, $R(A)$, is the column space of A. The null space of A, $N(A)$, is the collection of m-vectors x such that $Ax = \bar{0}$. It can be shown that

(1) $\dim R(A) + \dim N(A) = m$

Thus,

$$\text{rk } A + \dim N(A) = m$$

The reason for the name of $R(A)$ is clear since for any m-vector x, $Ax \epsilon$ column space of A.

We say that a vector space V is the direct sum of subspaces W_1 and W_2, written $W_1 \oplus W_2$, if every vector $x \epsilon V$ can be written uniquely as

$$x = x_1 + x_2 \text{ with } x_1 \epsilon W_1,\ x_2 \epsilon W_2.$$

If W is a subspace of V, the set of vectors orthogonal to W is a subspace of V called the orthogonal complement of W, denoted by $W^{\perp}$. It can be proved that $V = W \oplus W^{\perp}$. If a vector space V is the direct sum of W_1 and W_2, then $W_1 \cap W_2 = \{\bar{0}\}$.

If the vector $x \epsilon V$ is written in its unique form $x = x_1 + x_2$ with $x_1 \epsilon W$, $x_2 \epsilon W^{\perp}$, then the vector x_1 is called the orthogonal projection, or, more simply, the projection of x onto W. We write

$$x_1 = \text{proj}\ (x;\ W).$$

<u>Theorem I.3</u> Let A be nxm. Then $N(A'A) = N(A)$, and hence $\dim N(A'A) = \dim N(A)$.

Proof: We first show that $N(A'A) \subset N(A)$. Let $x \epsilon N(A'A)$. Then $A'Ax = \bar{0} \Rightarrow x'A'Ax = \bar{0} \Rightarrow (Ax,\ Ax) = 0 \Rightarrow Ax = 0 \Rightarrow x \epsilon N(A)$.

Conversely, let $Ax = 0$, then $A'Ax = 0$, and hence $x \epsilon N(A'A)$. Q.E.D.

<u>Corollary I.1</u>. If A is nxm, then $\text{rk } A = \text{rk } A'A = \text{rk } AA'$.

<u>Proof</u>: By (1),

$$m = \text{rk}(A'A) + \dim N(A'A) = \text{rk } A + \dim N(A) \text{ and from}$$

Theorem I.3, $\text{rk } A'A = \text{rk } A$.

Now interchanging A and A' in this result, we obtain rk AA' = rk A' = rk A. Q.E.D.

Let us remark, that although our development is confined to vector spaces over the real numbers, the analogous development for complex vector spaces requires only the substitution of A*, the conjugate transpose of A, for A', the ordinary transpose of A, in every statement involving A'.

II. GENERALIZED INVERSES

Perhaps the main application of matrix theory is the insight it gives us when we try to analyze a system of linear equations, say

$$
\begin{array}{c}
a_{11}\, x_1 + \ldots + a_{1\,m}\, x_m = y_1 \\
\vdots \\
a_{n_1}\, x_1 + \ldots + a_{n\,m}\, x_m = y_n
\end{array}
$$

It is well-known that this system can be rewritten as the matrix equation $Ax = y$, where $A = [a_{ij}]$,

$$x = \begin{bmatrix} x_1 \\ \\ x_n \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_1 \\ \\ y_n \end{bmatrix}$$

In the case where A is square and non-singular, a unique solution exists for every vector y. This is given by $x_0 = A^{-1}\, y$. In the case where A is square and singular or A is not square, the system may have a solution, or it may not have one.

The advantages of a generalized inverse, A^g, of an arbitrary matrix A are the following:

i) It always exists, and when the equation $Ax = y$ is consistent, i.e. has a solution, $x_0 = A^g y$ is one such solution.

ii) If $Ax = y$ is inconsistent, i.e. has no solution, the methods of working with a g.i. can be used to obtain a best approximate solution in the sense that we may find a vector x_o so that $\| Ax_o - y \|$ is as small as possible. That is, we may find a vector x_o so that

$$\| Ax_o - y \| = \inf_x \| Ax - y \|.$$

With this introduction, let us then make the

<u>Definition II.1</u> Let A be an n x m matrix. An m x n matrix A^g is called a g.i. of A if $AA^gA = A$.

<u>Theorem II.1</u> - (Bose). If A is n x m, a g.i. A^g of A always exists.

<u>Proof</u>: (Ref. 25) We know that we may row and column reduce A to obtain the matrix B where

$$B = \begin{bmatrix} I_r & \bar{0} \\ \bar{0} & \bar{0} \end{bmatrix}$$

The rk A = r, and the number of zeros in each block is such that B is n x m. This is equivalent to saying that there are non-singular matrices $\underset{n \times n}{P}$ and $\underset{m \times m}{Q}$ such that

(1) $PAQ = B$ which implies

$$A = P^{-1}BQ^{-1}$$

It is clear that if B^g is taken to be B', we have

(2) $BB^gB = B$.

We define $A^g = QB^gP$. Then,

$$AA^gA = P^{-1}\,BQ^{-1}\,QB^gPP^{-1}BQ^{-1}$$

$$= P^{-1}BB^gBQ^{-1} = P^{-1}BQ^{-1} = A. \quad \text{Q.E.D.}$$

<u>Corollary II.1</u>. If $PAQ = B$ where P and Q are non-singular matrices, and B^g is any g.i. of B, then QB^gP is a g.i. of A.

__Theorem II.2__ If A is nxm, PAQ = B as above, then every g.i. A^g of A is given by QB^gP where B^g is some g.i. of B.

__Proof:__ Let A^g be any g.i. of A. Then,

$$BQ^{-1}A^gP^{-1}B = PAQQ^{-1}A^gP^{-1}PAQ = PAA^gAQ = PAQ = B.$$

Thus $Q^{-1}A^gP^{-1}$ is a g.i. of B, and $A^g = Q(Q^{-1}A^gP^{-1})P$. Q.E.D.

We note that in general A^g is not unique. In fact, if PAQ = B, where

$$B = \begin{bmatrix} I_r & \bar{0} \\ \bar{0} & \bar{0} \end{bmatrix}$$

it is easily seen that we may define

$$B^g = \begin{bmatrix} I_r & U \\ V & W \end{bmatrix}$$

where U, V, W are arbitrary matrices of appropriate dimensions, and we still have $BB^gB = B$. Thus, in general, there may be many g.i.'s for a matrix.

The following theorem justifies the naming of a g.i.

__Theorem II.3__. If A is an n x n, non-singular matrix, then $A^g = A^{-1}$.

__Proof:__ Since A is non-singular, the only left identity for A is I_n. We know that AA^g is a left identity for A. Thus, $AA^g = I_n$, and $A^g = A^{-1}$ by left multiplication by A^{-1}. Q.E.D.

We should comment here that, in general $(AB)^g \neq B^gA^g$. However, if B and C are non-singular, then

$$(BAC)^g = C^{-1}A^gB^{-1}.$$

The next theorem, due to Penrose, gives the utility of the g.i. in solving a matrix equation.

<u>Theorem II.4</u>. Let A be n x m, B be p x q, C be n x q.

1) The matrix equation $A X B = C$ has a solution if and only if there are g.i.'s A^g of A and B^g of B so that $AA^gCB^gB = C$.

2) If $A X B = C$ is consistent, then we obtain the most general solution X by choosing A^g, B^g fixed, but arbitrary g.i.'s and setting

$$X = A^gCB^g + Y - A^gAYBB^g$$

where Y is an arbitrary m x p matrix.

<u>Proof</u>: 1) Suppose $A X B = C$ has the solution X_O, then if A^g and B^g are any g.i.'s of A and B,

$$C = AX_OB = AA^gA\, X_O\, BB^gB = AA^gCB^gB.$$

Conversely, if $AA^gCB^gB = C$ for some A^g and B^g, then A^gCB^g is a solution to $AXB = C$.

2) Suppose $AXB = C$ is consistent, and A^g, B^g are any g.i.'s of A and B. Then the matrix

$$X = A^gCB^g + Y - A^gA\, Y\, BB^g$$

is a solution for any m x p matrix Y. Conversely if X_0 is any solution, then X_0 is m x p, and $AX_0B = C$. Hence,

$$X_0 = A^gCB^g + X_0 - A^gCB^g$$

$$= A^gCB^g + X_0 - A^gA\,X_0\,BB^g. \qquad \text{Q.E.D.}$$

Corollary II.2 If A is n x m, x and y are vectors of appropriate dimensions, then $Ax = y$ is consistent if and only if there is a g.i. A^g so that $AA^gy = y$. When $Ax = y$ is consistent, the most general solution is given by

$$x = A^gy + (I_m - A^gA)\,z$$

where A^g is any g.i. of A, and z is an arbitrary m-vector.

We note that the general linear system

$$A_1XB_1 + A_2XB_2 + \ldots + A_rXB_r = C \tag{3}$$

may be rewritten as follows.

Suppose $\overset{n\times m}{A_k}$, $\overset{m\times p}{X}$, $\overset{p\times q}{B_k}$

Let i) $\overset{p\times q}{B_k} = \left[B_{ij}^{\,k}\right] = \left[B_1^{\,k} \mid \ldots \mid B_q^k\right]$

where B_i^k is the ith column of B^k.

ii) $X = [X_1 \ldots X_p]$, where X_i is the ith column of X.

iii) $\underset{mp \times 1}{X^*} = \begin{bmatrix} X_1 \\ \cdot \\ \cdot \\ \cdot \\ X_p \end{bmatrix}$

iv) $\underset{qpm \times 1}{X^{**}} = q\left\{\begin{bmatrix} X^* \\ \cdot \\ \cdot \\ \cdot \\ X^* \end{bmatrix}\right.$

v) $\underset{n \times pm}{D_{ik}} = \left[B_{1_i}^{\ k} A_k \mid \cdot \cdot \cdot \mid B_{pi}^{\ k} A_k \right]$

vi) $C = [C_1 \mid \cdot \cdot \mid C_q]$ where C_i is the ith column of C.

vii) $C^* = \begin{bmatrix} C_1 \\ \cdot \\ \cdot \\ \cdot \\ C_q \end{bmatrix}$

viii) $\underset{nq \times qpm}{H_k} = \begin{bmatrix} D_{1k} & 0 & \cdot & \cdot & 0 \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & D_{qk} \end{bmatrix}$

Then $\sum_{k=1}^{r} A_k X B_k = C$ may be rewritten $\left| \sum_{k=1}^{r} H_k \right| X^{**} = C^*$.

Thus (3) may be analyzed as above.

III. MINIMIZATION OF SYSTEMS OF LINEAR EQUATIONS

Now that the reader has seen the preceding results, he might wonder what happens if the system $Ax = y$ is inconsistent. It turns out that through the use of generalized inverses, we may obtain vectors which are as "close" to solutions as possible in the following sense.

Definition III.1. A vector $\bar{x}$ is said to minimize the equation $Ax = y$ if

$$\|A\bar{x} - y\| = \inf_{x} \|Ax - y\|$$

We shall show that there is a unique vector x_0 of smallest norm which minimizes $Ax = y$, and we will call this the minimal solution vector or, more simply, the minimal vector for $Ax = y$. When referring to the minimization of $Ax = y$, we may use the more suggestive terminology of the minimization of $\|Ax - y\|$.

To begin this discussion, we prove a well-known theorem which dates back to the time of Gauss.

Theorem III.1 The following conditions on a vector $\bar{x}$ are equivalent:

1) $\bar{x}$ minimizes $Ax=y$.

2) $A\bar{x} = \text{proj}\,(y;\, R(A))$.

3) $\bar{x}$ satisfies $A'Ax = A'y$.

Proof: We shall prove 1)$\Leftrightarrow$2) and then 2)$\Leftrightarrow$3). We need the following lemmas:

Lemma III.1. Let A be an $n \times m$ matrix then

a) $N(A)^{\perp} = R(A')$

b) $N(A') = R(A)^{\perp}$

Proof of Lemma III.1. Since a) follows from b) and the facts that $(R(A)^{\perp})^{\perp} = R(A)$, and $(A')' = A$, it is sufficient to prove b).

Let $x \in N(A')$. We show that $x \in R(A)^{\perp}$. That is, we must show if $t \in R(A)$, $(x,t) = 0$. We know that there is a y so that

$$t = Ay. \quad \text{Thus } (x,t) = (x,Ay) = y' A'x = y'0 = 0.$$

Hence $N(A') \subset R(A)^{\perp}$.

Now we show that if $x \notin N(A')$, then the vector $y = A(A'x) \in R(A)$, and $(y, x) \neq 0$. Indeed, if $A'x \neq \bar{0}$, then $(A'x, A'x) \neq 0$. But

$$(A'x, A'x) = x'AA'x = x'y = (y,x).$$

Hence $(y,x) \neq 0$. Q.E.D.

Lemma III.2. Let A be n x m, and $Ax = y$ where x is an m-vector, y an n-vector. Then

$$\alpha = \inf_x \| Ax-y \|^2 = \| y_1 \|^2$$

where

$$y_1 = \text{proj}\,(y; R(A)^{\perp}).$$

Proof of lemma III.2. Since y is an n-vector, we may write

$$y = y_0 + y_1 \text{ with } y_0 \in R(A),\ y_1 \in R(A)^{\perp}.$$

Then

$$\| Ax-y \|^2 = \| Ax-y_0 \|^2 + \| y_1 \|^2 \geq \| y_1 \|^2 .$$

Since $y_0 \in R(A)$, we conclude that

$$\alpha = \|y_1\|^2. \quad \text{Q.E.D.}$$

<u>Corollary III.1</u> $\|Ax-y\|^2 = \alpha$ if and only if

$$Ax = y_0 = \text{proj } (y;\ R(A)).$$

Now to the proof of theorem III.1: 1) $\Leftrightarrow$ 2). This follows easily from Corollary III.1.

2) $\Rightarrow$ 3). Suppose 2) is true. Let $A\bar{x} = y_0$ then

$$A'y = A'y_0 + A'y_1 = A'y_0$$

since $y_1 \in R(A)^{\perp} = N(A')$. Thus $A'y = A'A\bar{x}$ and 3) is true.

Conversely, let $\bar{x}$ be such that $A'A\bar{x} = A'y$ then

$$A\bar{x} \in R(A) \text{ and } y_0 \in R(A)$$

$$\Rightarrow A\bar{x} - y_0 \in R(A).$$

But

$$\bar{0} = A'A\bar{x} - A'y = A'A\bar{x} - A'y_0 = A'(A\bar{x} - y_0)$$

so

$$A\bar{x} - y_0 \in N(A') = R(A)^{\perp}$$

$$\Rightarrow A\bar{x} - y_0 = \bar{0} \Rightarrow A\bar{x} = y_0 \quad \text{and hence}$$

$\bar{x}$ minimizes $\|Ax-y\|$. Q.E.D.

The advantage of this result is that it enables us to translate the problem of minimizing $\|Ax-y\|$ to the easier problem of finding a solution to the equation $A'Ax = A'y$. Such a solution always exists since $A'y \in R(A') = R(A'A)$. Later we shall be interested in finding the unique solution x_0 of $A'Ax = A'y$

which has smallest norm, i.e., the minimal vector for $Ax = y$, but now we content ourselves with finding any x such that $A'Ax = A'y$.

According to what has been said (Corollary II.2), a solution to $A'Ax = A'y$ is given by $\bar{x} = (A'A)^g A'y$ where $(A'A)^g$ is any g.i. of $A'A$. Since the equations $A'Ax = A'y$ are called the normal equations for $Ax = y$, we are led to the following.

Definition III.2. If A is an $n \times m$ matrix, then the $m \times n$ matrix $A^n = (A'A)^g A'$ where $(A'A)^g$ is any g.i. of $A'A$ is called a normalized generalized inverse of A.

This structure was first studied by Zelen, who used the term weak generalized inverse.

We observe first that $AA^nA = A$, so A^n is a g.i. of A. In fact, A^n satisfies many more properties which we shall study later, and which we shall use to give an equivalent definition of A^n. Let us note, in this terminology, that the vector $\bar{x} = A^n y$ minimizes the equation $Ax = y$ where A^n is any normalized g.i. of A.

IV. PSEUDO-INVERSES

Let us study the normalized g.i. A^n. We will show that if A^n is a normalized g.i. with the property that $(A^nA)' = A^nA$, then the minimal vector of $Ax = y$ is $x_0 = A^ny$. But, before we do this, let us look at things from a slightly different point of view.

The following lemma and corollary which are due to Bose, are found in the paper by Rohde (Ref. 25).

Lemma IV.1.

Let X' be a $p \times n$ matrix. There exists a $p \times n$ matrix Y such that

1) $(X'X)Y = X'$

2) XY is unique in the sense that if $X'XY_1 = X'$ and $X'XY_2 = X'$ $\quad XY_1 = XY_2$.

3) $(XY)' = XY$

4) $(XY)^2 = XY$.

Proof

1) To prove the existence of Y, all we need observe is that every column of X' is in the column space of $X'X$. Hence, there is a Y such that $X'XY = X'$.

2) Suppose $X'XY_1 = X'$ and $X'XY_2 = X'$. We will show that for $i = 1, \ldots, n$, the ith columns of XY_1 and XY_2 are equal.

Let Y_{1i} be the ith column of Y_1, and Y_{2i} be the ith column of Y_2.

Then $X'XY_{1i} = X'XY_{2i}$

$\Rightarrow X'X(Y_{1i} - Y_{2i}) = \bar{0}$

$\Rightarrow (Y_{1i} - Y_{2i})'X'X(Y_{1i} - Y_{2i}) = \bar{0}$

$\Rightarrow$ the inner product $(X(Y_{1i} - Y_{2i}), X(Y_{1i} - Y_{2i})) = 0.$

$\Rightarrow XY_{1i} = X Y_{2i}$. Q.E.D.

3) $(XY)' = Y'X' = Y'(X'XY) = (Y'X'X)Y$

$= (X'XY)'Y = (X')'Y = XY.$

4) $(XY)^2 = (XY)(XY) = (XY)'XY$

$= Y'X'XY = Y'X' = (XY)' = XY.$ Q.E.D.

Corollary IV.1

The matrix $A(A'A)^gA'$, where $(A'A)^g$ is any g.i. of $A'A$, is uniquely determined, symmetric and idempotent.

Proof

The equation $A'AX = A'$ has a solution given by $X = (A'A)^gA'$. Hence $AX = A(A'A)^gA'$ has the desired properties, by lemma IV.1.

Theorem IV.1.(Rohde)

The matrix A^n is a normalized g.i. of A if and only if A^n satisfies the following:

5) $AA^nA = A$

6) $A^nAA^n = A^n$

7) $(AA^n)' = AA^n$.

Proof

Necessity.

If A^n is an n.g.i. of A, there is a matrix $(A'A)^g$ such that

$$A^n = (A'A)^gA'.$$

Then $AA^nA = A(A'A)^gA'A = A$, since $(A'A)^gA'A$ is a right identity for $A'A$, and hence is a right identity for A.

Similarly,

$$A^nAA^n = (A'A)^gA'A\ (A'A)^gA' = (A'A)^gA'.$$

Thus properties 5) and 6) are satisfied. Property 7) is satisfied by Corollary IV.1.

Sufficiency.

From properties 6) and 7), we have that

row space $A^n \subset$ row space AA^n

= column space $AA^n \subset$ column space A

= row space A'.

Thus, there is a matrix X such that $A^n = XA'$. We show that X is a g.i. of A'A.

Indeed, $A'AXA'A = A'AA^nA = A'A$ by property 5) Q.E.D.

The characterization of normalized g.i.'s given in the above theorem is what is usually used to define these structures. Pursuing this type of reasoning, let us, for an arbitrary n x m matrix A, consider the following equations:

8) $AXA = A$

9) $XAX = X$

10) $(AX)' = AX$

11) $(XA)' = XA$

We have shown that 8), 9), 10) have a solution for any A, and we shall presently show that in fact these four equations have a unique solution for any A. But, before we do this, let us just mention that in the literature so far there have been four types of g.i.'s studied. These are obtained as follows:

For an arbitrary matrix A, a matrix X is called

a generalized inverse (g-inverse) if it satisfies 8).

a reflexive generalized inverse (r-inverse) if it satisfies 8), 9).

a normalized generalized inverse (n-inverse) if it satisfies 8), 9), and 10).

a pseudo-inverse (p-inverse) if it satisfies 8), 9), 10), and 11).

Our main purpose is the study of the p-inverse which has the particularly nice property that it is unique. However, to facilitate some later proofs, we shall give the following two results due to Rohde, who studied in some detail the properties of the four types of g.i.'s (Ref. 25).

Theorem IV.2.

For any n x m matrix A, and any g.i. A^g of A, rank $A^g \geq$ rank A, and rank A^gA = rank AA^g = rank A.

Proof: By theorem I.2.

$$\text{rk } A \leq \text{rk } A^gA \leq \text{rk } A^g;$$

further,

$$\text{rk} A \geq \text{rk } AA^g \geq \text{rk } AA^gA = \text{rk } A,$$

and

$$\text{rk } A \geq \text{rk } A^gA \geq \text{rk } AA^gA = \text{rk } A. \quad \text{Q.E.D.}$$

Theorem IV.3

The matrix A^g is an r-inverse of A if and only if rk A^g = rk A.

Proof:

If A^g is an r-inverse of A, then A is a g.i. of A^g. Hence

$$\text{rk } A \leq \text{rk } A^g \leq \text{rk } A$$

$$\text{rk } A = \text{rk } A^g.$$

Now suppose rk A^g = rk A.

By Theorem II.2 and the remarks which follow it, we know that

$$A^g = QB^gP \text{ where } PAQ = B = \begin{bmatrix} I_r & \bar{0} \\ \bar{0} & \bar{0} \end{bmatrix}$$

and $B^g = \begin{bmatrix} I_r & U \\ V & W \end{bmatrix}$, where U, V, W are arbitrary matrices of appropriate dimensions.

We have that rk A^g = rk A = rk B and rk A^g = rk B^g since multiplication by non-singular matrices preserves rank.

Thus rk B = rk B^g.

If we show that this implies that $B^gBB^g = B^g$, we have

$$A^gAA^g = QB^gPP^{-1}QB^gP = QB^gBB^gP$$

$$= QB^gP = A^g$$

which is our desired result.

Thus we must show $B^gBB^g = B^g$.

We have

$$\operatorname{rk} B^g = \operatorname{rk} \begin{bmatrix} I_r & \bar{0} \\ -V & I_{m-r} \end{bmatrix} \begin{bmatrix} I_r & U \\ V & W \end{bmatrix} \begin{bmatrix} I_r & -U \\ \bar{0} & I_{n-r} \end{bmatrix}$$

$$= \operatorname{rk} \begin{bmatrix} I_r & U \\ \bar{0} & W-VU \end{bmatrix} \begin{bmatrix} I_r & -U \\ \bar{0} & I_{n-r} \end{bmatrix}$$

$$= \operatorname{rk} \begin{bmatrix} I_r & \bar{0} \\ \bar{0} & W-VU \end{bmatrix} \geq \operatorname{rk} \begin{bmatrix} I_r & \bar{0} \\ \bar{0} & \bar{0} \end{bmatrix}$$

$$= \operatorname{rk} B$$

Hence,

$$\operatorname{rk} B^g = \operatorname{rk} B \Rightarrow W-VU = 0$$

$$\Rightarrow W = VU \Rightarrow B^g = \begin{bmatrix} I_r & U \\ V & VU \end{bmatrix}$$

Thus

$$B^g B B^g = \begin{bmatrix} I_r & U \\ V & VU \end{bmatrix} \begin{bmatrix} I_r & \bar{0} \\ \bar{0} & \bar{0} \end{bmatrix} \begin{bmatrix} I_r & U \\ V & VU \end{bmatrix}$$

$$= \begin{bmatrix} I_r & \bar{0} \\ V & \bar{0} \end{bmatrix} \begin{bmatrix} I_r & U \\ V & VU \end{bmatrix} = \begin{bmatrix} I_r & U \\ V & VU \end{bmatrix}$$

$$= B^g \quad \text{Q.E.D.}$$

Henceforth in this paper we shall be concerned with pseudo-inverses. We begin by giving the basic theorem of Penrose (Ref. 20) concerning the existence and uniqueness of p-inverses.

Theorem IV.3.

For any n x m matrix A, there is a unique m x n matrix A^+, called the pseudo-inverse (p-inverse) of A, satisfying:

12) $AA^+A = A$

13) $A^+AA^+ = A^+$

14) $(AA^+)' = AA^+$

15) $(A^+A)' = A^+A$

Proof

We first note that $AA' = 0 \Leftrightarrow A = 0$ and then that

16) $BAA' = CAA' \Rightarrow BA = CA$

and

17) $A'AB = A'AC \Rightarrow AB = AC$

These follow respectively from

18) $(BAA' - CAA')(B-C)' = (BA - CA)(BA - CA)'$

and

19) $(B - C)'(A'AB - A'AC) = (AB - AC)'(AB - AC)$

Now, since row space AA' = row space A', there is a matrix W such that $WAA' = A'$. Similarly, since column space A' = column space $A'A$, there is a matrix Y such that $A'AY = A'$.

Let us define $A^+ = WAY$.

Since $AWAA' = AA'$ and $A'AYA \doteq A'A$, we have that $AWA = A$ and $AYA = A$. Thus $AA^+A = AWAYA = AYA = A$ and $A^+AA^+ = WAYAWAY = WAWAY = WAY = A^+$. Hence 12) and 13) are satisfied.

Now,

$$(AY)' = Y'A' = Y'A'AY = (AY)'AY$$

which shows that $(AY)'$ is symmetric and hence $AY = (AY)'$.

Similarly, $WA = (WA)'$.

Thus

$$AA^+ = AWAY = AY = (AY)' = (AA^+)'$$

and

$$A^+A = WAYA = WA = (WA)' = (A^+A)'$$

Hence $A^+ = WAY$ is a p-inverse of A.

Uniqueness:

Let us note that for any p-inverse A^+ of A, $\text{rk } A^+ = \text{rk } A$.

Further, we have

$$\text{row space } A^+ \subset \text{row space } AA^+$$

$$= \text{column space } AA^+ \subset \text{column space } A.$$

$$= \text{row space } A'$$

$$\Rightarrow \text{row space } A^+ = \text{row space } A'.$$

Similarly, column space A^+ = column space A'. Thus $A'AA^+ = A'$, and $A^+AA' = A'$.

Now, let G and X be p-inverses of A.

Then

$$GAA' = A' \Rightarrow GAX = X \text{ and}$$

$$A'AX = A' \Rightarrow GAX = G \quad \text{Q.E.D.}$$

The following properties of the p-inverse were obtained by Penrose.

<u>Theorem IV.4.</u>

Let A be a matrix. Then

20) $(A^+)^+ = A$

21) $(A')^+ = (A^+)'$

22) $A^+ = A^{-1}$ if A is non-singular

23) $(\lambda A)^+ = \lambda^+ A^+$, λ a real number,

$$\text{where } \lambda^+ = \begin{cases} 0 \text{ if } \lambda = 0 \\ \frac{1}{\lambda} \text{ if } \lambda \neq 0. \end{cases}$$

24) $(A'A)^+ = A^+A^{+}{}'$

25) If U and V are orthogonal matrices, then $(UAV)^+ = V'A^+U'$.

26) If $A = \Sigma A_i$ where $A_i A_j' = 0$ and $A_i{}' A_j = 0$ for $i \neq j$, then $A^+ = \sum_i A_i{}^+$.

27) $A^+ = (A'A)^+A' = A'(AA')^+$

28) If A is symmetric and $PA P' = D = \text{diag}\ \{\lambda_1,..,\lambda_n,0,..0\}$, then $A^+ = PD^+P'$ where $D^+ = \text{diag}\ \{\lambda_1^{-1},..,\lambda_n^{-1}, 0,..,0\}$.

29) A^+A, AA^+, $I-A^+A$, $I-AA^+$ are all symmetric, indempotent matrices.

30) If A is normal, i.e. $AA' = A'A$, then $AA^+ = A^+A$ and $(A^m)^+ = (A^+)^m$ for m a positive integer.

31) A, $A'A$, A^+, A^+A, AA^+ all have rank equal to Trace A^+A.

Proof

We shall sketch parts of the proof.

Property 24)

We have $A'AA^+A^{+}{}'A'A$

$= A'A^{+}{}'A'A$ (See proof of Theorem IV.3.).

$= A'A$. The other properties are similar.

Property 27)

We have $A(A'A)^+A'A = A$ since $(A'A)^+ A'A$ is a right identity for $A'A$, and hence is a right identity for A.

Property 29)

Suppose $AA' = A'A$. Then

$$AA^+ = (AA^+)' = A^{+\prime}A' = A^{+\prime}A^+AA'$$
$$= (AA')^+AA' = (A'A)^+ A'A = A^+A.$$

Property 30)

We know that $\mathrm{rk}\, A = \mathrm{rk}\, A'A = \mathrm{rk}\, A^+$ by Theorems I.3 and IV.3.

Now

$$\mathrm{rk}\, A = \mathrm{rk}\, AA^+A \leq \mathrm{rk}\, A^+A = \mathrm{rk}\, A$$

so $\mathrm{rk}\, A = \mathrm{rk}\, A^+A$, and similarly $\mathrm{rk}\, A = \mathrm{rk}\, AA^+$. Since A^+A is symmetric and indempotent, there is an orthogonal matrix P such that

$$PA^+AP' = \mathrm{diag}\ \{1, 1, .., 1, 0, .., 0\}.$$

Then

$$\mathrm{tr}\, A^+A = \mathrm{tr}\, PA^+AP' = \mathrm{rk}\, A^+A. \qquad \text{Q.E.D.}$$

We observe that, in general, $(AB)^+ \neq B^+A^+$. However, Cline (Ref. 5) has shown that we may find matrices B_1 and A_1 such that

$$AB = A_1B_1 \quad \text{and} \quad (AB)^+ = B_1{}^+A_1{}^+$$

In fact $B_1 = A^+AB$, and $A_1 = AB_1B_1{}^+$.

Now that we have the p-inverse, let us return to our problem of finding the minimal solution vector of a matrix equation $Ax = y$ where A is n x m. Recall that we have shown that a vector x_o minimizes the equation $Ax = y$ if and only if $A'Ax_o = A'y$. We have also shown that $A'Ax_o = A'y$ if and only if $(A'A)^gA'y = A'y$, and that if $x_0 = (A'A)^g A'y$, then x_0 minimizes $Ax = y$, where $(A'A)^g$ is an arbitrary g.i. of $A'A$.

We thus know that if $x_0 = A^n y$, for some n-inverse A^n of A then x_0 minimizes $Ax = y$. We now ask what further conditions we must put on x_0 so that it will be the minimal vector for $Ax = y$. We shall show presently that a necessary and sufficient condition for x_0 to be the minimal vector is that it minimize $Ax = y$ and belong to $R(A')$. We will also show that there is only one minimal vector for $Ax = y$. For the time being, let us assume we know these results. We then could write any minimizing vector $\bar{x}$ as $\bar{x}_1 + x_2$ where $x_1 \in R(A')$ and $x_2 \in R(A')^{\perp}$. Then, since $Ax_2 = 0$, $(R(A')^{\perp} = N(A))$, we would have that $A\bar{x} = y$, and $x_1 \in R(A')$. Thus we would have that $A\bar{x} = A\bar{x}_1$, which would mean that x_1 also minimized $Ax = y$, and $x_1 \in R(A')$. But we know that we can obtain a minimizing vector $\bar{x}$ by setting $\bar{x} = A^n y$ where A^n is any n-inverse of A. From what we have said, if $A^n y$ were in $R(A')$, $A^n y$ would be precisely the minimal vector for $Ax = y$.

Now it seems reasonable to ask just what restrictions on A^n we need to insure that $A^n y \in R(A')$. This we state as the following theorem which is another form of a result stated by Albert (Ref. 2).

Theorem IV.5.

Let A be an n x m matrix; A^n an n-inverse of A. Then if $(A^n A)' = A^n A$, i.e., if A^n is the p-inverse of A, $A^n y = A^+ y \in R(A')$ for all y. Conversely, if an n-inverse A^n is such that $A^n y \in R(A')$ for all y, then $A^n = A^+$.

Proof

If $A^n = A^+$, then $A^+ y \in R(A^{+\prime}) = R(A')$

Conversely, if $A^n y \in R(A')$ for all y, then we must show that $A^n = A^+$. In other words, we must show that

A^+ is the only matrix M satisfying:

32) row space $M \subset$ row space A'.

33) column space $M \subset$ column space A'.

34) AM is a left identity for A.

35) MA is a right identity for A.

We first prove that AA^+ is the only left identity for A with rows in the row space of A', and A^+A is the only right identity for A with columns in the column space of A'.

Indeed, suppose that B is a left identity of A of the form $X A'$.

Then

$$XA'A - AA^+A = 0$$

$$\Rightarrow XA'A - A(A'A)^+A'A = 0 \text{ by 27)}$$

$$\Rightarrow XA' = A(A'A)^+A' \text{ by 16) and the fact that } (A')' = A.$$

Thus

$$B = XA' = AA^+.$$

Similarly, A^+A is the only right identity for A with columns in the column space of A'.

Now, to complete the proof of Theorem IV.5., suppose that A_1 is any matrix satisfying 32) - 35).

Then

$$A_1 = A_1AA^+ = A^+AA^+ = A^+. \quad \text{Q.E.D.}$$

The basic idea of the last proof is found in a paper by Greville (Ref. 13).

We now prove the theorem which will pick up all the loose ends we have left.

<u>Theorem IV.6</u>

36) A minimizing vector x_0 for the matrix equation $Ax = y$

is the minimal vector for $Ax = y$ if and only if $x_0 \in R(A')$.

37) The minimal vector x_0 is unique, and is given by $x_0 = A^+y$.

Proof

Suppose x_0 is the minimal vector for $Ax = y$. Then x_0 has the unique decomposition $x_0 = x_1 + x_2$ where $x_1 = \text{proj}\ (x_0; R(A'))$ and $x_2 = \text{proj}\ (x_0\ ;\ R(A')^{\perp}\ )$.

Since $x_2 \in R(A')^{\perp} = N(A)$, we have by Theorem III.1, $Ax_0 = Ax_1 = \text{proj}\ (y; R(A))$. Thus x_1 also minimizes $Ax = y$. But

$$\|x_0\|^2 = \|x_1\|^2 + \|x_2\|^2 \geq \|x_1\|^2$$

$$\Rightarrow \|x_0\|^2 = \|x_1\|^2 \Rightarrow x_2 = \bar{0} \Rightarrow x_0 = x_1$$

Hence

$$x_0 \in R(A').$$

Conversely, let x_0 be a minimizing vector for $Ax = y$ which is in the range of A'. Let $\bar{x}$ be any minimizing vector for $Ax = y$. We shall show that $\|\bar{x}\| \geq \|x_0\|$, and, in fact, $x_0 = \text{proj}\ (\bar{x}; R(A'))$. Write $\bar{x} = t_1 + t_2$ with $t_1 \in R(A')$, $t_2 \in R(A')^{\perp} = N(A)$.

Then

$$A\bar{x} = At_1 = \text{proj}\ (y; R(A)) = Ax_0$$
$$\Rightarrow x_0 - t_1 \in N(A) = R(A')^{\perp}$$

But

$$x_0 - t_1 \in R(A') \Rightarrow x_0 = t_1.$$

Now for the uniqueness, if x_0, t_0 are two minimal vectors for $Ax = y$, the $x_0 = \text{proj}\ (t_0; R(A')) = t_0$.

Since A^+y minimizes $Ax = y$, and $A^+y \in R(A')$, A^+y is the minimal vector for $Ax = y$. Q.E.D.

We shall summarize our results in the following theorems.

Theorem IV.7

Let A be an n x m matrix with real entries. Let x be an m-vector; y, an n-vector; $\alpha = \inf_x \| Ax - y\|$. Then the minimal vector x_o for $Ax = y$ is $x_o = A^+y$ where A^+ is the p-inverse of A. This vector satisfies

38) $\|Ax_o - y\| = \alpha$, and if $\bar{x}$ is such that $\bar{x} \neq x_o$ and $\|A\bar{x} - y\| = \alpha$, then $\|\bar{x}\| > \|x_o\|$.

39) $x_o \in R(A')$.

40) $Ax_o = \text{proj } (y; R(A))$.

These results are a slight reformulation of those found in Albert (Ref. 2). The following theorem, which is to be found in Albert, summarizes the main applications of the p-inverse.

Theorem IV.8

41) Let A be an n x m matrix with column vectors $A_1, \ldots, A_m$. Then, if $L(A_1, \ldots, A_m) = R(A)$ is the space spanned by these vectors, $y_o = \text{proj } (y; L(A_1, \ldots, A_m)) = AA^+y$.

42) $y_1 = \text{proj } (y; N(A')) = (I-AA^+)y$

43) $\bar{x}$ minimizes $\|Ax-y\|$ if and only if there is a z such that

$$\bar{x} = A^+y + (I-A^+A)Z.$$

In closing this section, we observe that there are at least three alternative methods of defining the p-inverse A^+ of a matrix A. The first two methods we have already mentioned, and we state now as

Theorem IV.9.

44) The p-inverse A^+ of A is the unique matrix such that for any m-vector y, $A^+y \in R(A')$, and A^+y minimizes $Ax = y$.

45) The p-inverse A^+ is the unique matrix satisfying 32) - 35).

Albert (Ref. 2) uses another equivalent definition which we give as

Theorem IV.10.

For any n x m matrix A,

$$A^+ = \lim_{\epsilon \to 0} (A'A + \epsilon I)^{-1} A'$$

$$= \lim_{\epsilon \to 0} A'(AA' + \epsilon I)^{-1}$$

where $A'A + \epsilon I$ will always be invertible if $|\epsilon|$ is less than the absolute value of the smallest non-zero characteristic value of $A'A$.

V. COMPUTATION OF THE PSEUDO-INVERSE

Having discussed at length the geometric applications of the p-inverse, it seems desirable to have at hand an economical method for its computation. In view of property 27) of section IV, it suffices to find the p-inverse of matrices of the form $A'A$. The following theorem shows that it suffices, in fact, to find an arbitrary g.i. of a matrix of the form $A'A$.

Theorem V.1.

If A is symmetric, then $A^+ = A[(A^2)^g A]^2$ where $(A^2)^g$ is an arbitrary g.i. of A^2.

Proof

This is a straightforward application of Penrose's method for the computation of the p-inverse.

We solve 1) $WA^2 = A$ and

2) $A^2 Y = A$.

A solution of 1) is $W_o = A(A^2)^g$, and a solution of 2) is $Y_o = (A^2)^g A$.

Then

$$A^+ = W_o \, A \, Y_o = A(A^2)^g A(A^2)^g A$$
$$= A[(A^2)^g A]^2. \quad \text{Q.E.D.}$$

Now suppose $A = H'H$. Then to find A^+, all we need to find is $(A^2)^g$. Since $A^2 = (H'H)\,(H'H) = (H'H)'(H'H)$, we have reduced the problem to that of finding the g.i. of a matrix of the form $H'H$.

Let us remark that in the proof of the existence of a g.i. A^g of an arbitrary matrix, we showed how to compute one. Our method involved a row reduction, i.e. pre-multiplication by a non-singular matrix P, and a column reduction, i.e. post multiplication by a non-singular matrix Q.

Rao has shown (Ref. 23) that when A is of the form H'H, we may row-reduce A by pre-multiplication by a non-singular matrix P, and that P is a g.i. of A, i.e. APA = A. Applying this to the above theorem, we see that finding $(A'A)^g$ involves only ordinary row reduction. To prove Rao's result we need some additional terminology.

Definition

We say that the n x n matrix A has the row-reduced zero property (A has r.r.z.p.) if its row reduced echelon form has the property that when its ith diagonal element is zero, its ith row is composed only of zeros.

Lemma V.1. (Rao)

Let A be an n x n matrix with r.r.z.p. Let R be its row reduced echelon form. Let P be a non-singular matrix such that PA = R. Then

3) R is idempotent; i.e. $R^2 = R$.

4) AR = R

5) APA = A, and hence P is a g.i. of A.

6) A necessary and sufficient condition that Ax = y be consistent is that if the r_1th, r_2th, ... coordinates of Py must be null.

7) A general solution of Ax = y is Py + (I - PA)z where z is arbitrary.

Proof

3) Let $R = [r_{ij}]$ Let $R^2 = R \cdot R = C = [c_{ij}]$. Then

$$c_{ij} = \sum_{k=1}^{n} r_{ik} r_{kj}$$

We have three cases

Case 1: $i > j$

Then

$$c_{ij} = \sum_{k=1}^{j} r_{ik}\, r_{kj} + \sum_{k=j+1}^{n} r_{ik} r_{kj}$$

$$= 0 = r_{ij}.$$

Case 2: Let $i = j$ be fixed. Then either $r_{jj} = 1$ or $r_{jj} = 0$. If $r_{jj} = 1$, then $r_{\ell j} = 0$ for $\ell \neq j$, i.e. for all other entries in the jth column of R.

Then the jth column of C is the same as the jth column of R. If $r_{ij} = 0$, the jth row of R is composed only of zeros. Hence, the same is true of the jth row of C.

Thus $r_{ij} = c_{ij}$ for all $i \geq j$.

Case 3: Now let r_{ij} be fixed where $i < j$. Then $r_{ij} \neq 0$ if and only if $r_{ii} = 1$ and $r_{jj} = 0$. This implies that $c_{ij} = r_{ij}$. Thus $R^2 = R$.

4) $PA = R \Rightarrow A = P^{-1}R = P^{-1} R \cdot R = AR$.

5) We have $APA = AR = A$.

6) See Hoffman and Kunze (Ref. 17), Chapter 1.

7) This follows from Corollary II.2. Q.E.D.

Theorem V.2. (Rohde)

If $A = H'H$ where H is any matrix, then A has the row reduced zero property and hence the results of lemma V.1. are true for A.

Proof

We prove this by induction on the dimension of A.

Let $A = H'H$ be an $n \times n$ matrix, and assume that the result is true for all matrices B of the form $B = S'S$ where dimension $B < n$. We may begin to row reduce A. Suppose that we have completed row reduction of the first i-1 rows, and we wish to continue with the reduction of the $i\underline{th}$ row. Let us see what we have done.

We started with A which we can write as follows:

$$A = \begin{bmatrix} H_1' \\ x_1' \\ H_2' \end{bmatrix} [H_2 x_i H_2] = \begin{bmatrix} H_1'H_1 & H_1'x_i & H_2'H_2 \\ x_i'H_1 & x_i'x_i & x_i'H_2 \\ H_2'H_1 & H_2'x_i & H_2'H_2 \end{bmatrix}$$

where H_1 is the matrix of the first i-1 columns of A, x_i is the $i\underline{th}$ column of A, and H_2 is the matrix of the remaining columns of A.

We first observe that the (i-1) x (i-1) matrix $H_1'H_1$ satisfies our inductive hypothesis. Thus, if we row reduce $H_1'H_1$, we do this by pre-multiplication by a non-singular matrix $(H_1'H_1)^g$. With this comment it should be clear that the non-singular matrix which has row reduced the first i-1 rows of A is

$$\begin{bmatrix} (H_1'H_1)^g & 0 & \bar{0} \\ 0 & 1 & 0 \\ \bar{0} & 0 & I \end{bmatrix}$$

Now that we wish to work with the $i\underline{th}$ row of A, let us do this in detail. We have put A in the form

$$\begin{bmatrix} (H_1'H_1)^g & 0 & \bar{0} \\ 0 & 1 & 0 \\ \bar{0} & 0 & I \end{bmatrix} \begin{bmatrix} H_1'H_1 & H_1'x_i & H_1'H_2 \\ x_i'H_1 & x_i'x_i & x_i'H_2 \\ H_2'H_1 & H_2'x_i & H_2'H_2 \end{bmatrix}$$

$$= \begin{bmatrix} R_1 & (H_1'H_1)^g H_2'x_i & (H_1'H_1)^g H_1'H_2 \\ x_i'H_1 & x_i'x_i & x_i'H_2 \\ H_2'H_1 & H_2'x_i & H_2'H_2 \end{bmatrix}$$

where R_1 is the row reduced form of $H_1'H_1$.

To reduce the $\underline{ith}$ row of this matrix, we pre-multiply by

$$\begin{bmatrix} I & 0 & \bar{0} \\ -x_i{}'H_1 & 1 & 0 \\ \bar{0} & 0 & I \end{bmatrix}$$

to obtain

$$\begin{bmatrix} R_1 & (H_1{}'H_1)^g H_1{}'x_i & (H_1{}'H_1)^g H_1{}'H_2 \\ 0 & f_i & g_i \\ H_2{}'H_1 & H_2{}'x_i & H_2{}'H_2 \end{bmatrix}$$

where

$$f_i = x_i{}'x_i - x_i{}'H_1(H_1{}'H_1)^g H_1{}'x_i$$

and

$$g_i = x_i{}'H_2 - x_i{}'H_1(H_1{}'H_1)^g H_1{}'H_2$$

Now we assume $f_i = 0$, and we wish to show that $g_i = 0$. We have that

$$0 = f_i = x_i{}' \, [I - H_1(H_1{}'H_1)^g H_1{}'] \, x_i .$$

By Corollary IV.1., $H_1(H_1{}'H_1)^g H_1{}'$ is symmetric and idempotent, hence, so is

$$I-H_1(H_1{}'H_1)^g H_1{}'$$

Thus

$$0 = f_i = x_i{}'(I-H_1(H_1{}'H_1)^g H_1{}')' \, (I-H_1{}'(H_1{}'H_1)^g H_1{}'')x_i$$

$$\Rightarrow x_i{}' \, (I-H_1 \, (H_1{}'H_1)^g H_1{}') = 0$$

$$\Rightarrow x_i{}'(I-H_1(H_1{}'H_1)^g H_1{}')H_2 = 0$$

$$\Rightarrow 0 = x_i{}'H_2-x_i{}'H_1(H_1{}'H_1)^g H_1{}'H_2 = g_i$$

Since this is true for $i < n$, A has r.r.z.p. Q.E.D.

Let us recapitulate briefly. To find a g.i. of the $n \times n$ matrix $A = H'H$, we merely adjoin I_n to the right of A to get

$$[A|I_n]$$

Then we row reduce this matrix and get

$$[A^gA|A^g]$$

To consolidate our computational method, we have

Theorem V.3.

Let A be an arbitrary $n \times m$ matrix. Then

$$A^+ = (A'A)^+A' = A'(AA')^+$$

$$\text{and } (A'A)^+ = A'A\ [((A'A)^2)^gA'A]^2$$

$$\text{Thus } A^+ = A'A\ [((A'A)^2)^gA'A]^2A'$$

and

a similar formula holds involving AA'. This method involves six matrix multiplications and one row reduction.

So far, the simplest method we have seen for finding the g.i. of an arbitrary matrix involves a row reduction, column reduction, and two matrix multiplications (Theorem II.2.). The following method, due to Frame (Ref. 11), shows that we may find a g.i. of an $n \times m$ matrix A by little more than ordinary row reduction.

Definition

8) The distinguished columns of the $n \times m$ matrix A of rank s, are those linearly independent columns which are obtained by starting at the first column on the left, moving to the right, and deleting any column which is a linear combination of the columns preceding it.

9) A is said to have the rank factorization $A = BC$ where B is the submatrix of the distinguished columns of A in their natural

order, and C is the submatrix of the top s rows of the row reduced echelon form R of A.

We note that rk A = rk B = rk C, and that every matrix has a unique rank factorization.

Now we keep A an n x m matrix of rank s. Let L be an n x n non-singular matrix such that LA = R, where R is the row reduced echelon form of A. Let L_1 be the submatrix of the top s rows of L, and C be the submatrix of the top s rows of R. Let L_2 be the submatrix of the remaining rows of R. The other rows of C are composed only of zeros. The distinguished columns of C are those of I_s. If we let V be the submatrix of the other columns of C in their original order, then, writing the s x m partitioned matrix $[I_s|V]$, we see that we may interchange columns in $[I_s|V]$ to get back to C. Letting P be the matrix obtained by interchanging the appropriate columns of I_m, we have $C = [I_s|V]P$. Now, we may write $P = \left[\frac{P_1}{P_2}\right]$ where the s x m matrix P_1 consists of the distinguished columns of C, and the (m - s) x m matrix P_2 consists of the other rows of the identity. Because of the way we have chosen $P = \left[\frac{P_1}{P_2}\right]$ we have that

$$C = [I_s|V] \left[\frac{P_1}{P_2}\right] = P_1 + VP_2$$

and

$$P^{-1} = P' = [P_1'|P_2']$$

We assert that the n x s matrix $B = AP_1'$ consists exactly of the distinguished columns of A. Let us see why this is so. The m x s matrix P_1' has as its columns some of the columns of I_m. Since the non-zero columns of P_1 are the distinguished columns of C, they occur in the same positions as the distinguished columns of A.

That is, the jth column of P_1 is non-zero if and only if the jth column of A is a distinguished column of A. Now the i,jth entry of P_1 is non-zero if and only if the ith row of P_1 is $e_j' = [0,..,0,1_j,0,..0]$. When we multiply AP_1', the ith column of this product is the jth column of A, since the ith column of this product is just A times the ith column of P_1', which is $[0..0,1_j,0..0]'$.

Using the fact that $B = AP_1'$ is the matrix of the s distinguished columns of A, we see that A has the rank factorization $A = BC = AP_1'L_1A$.

We state this formally with another observation as

Theorem V.4.

Let A be an n x m matrix of rank s. Let L, C, L_1, L_2, P_1, P_2 be defined as above. Then the matrix $A^r = P_1'L_1$ is a reflexive g.i. of A; i.e. A^r satisfies $AA^rA = A$, and $A^rAA^r = A^r$.

Proof

By the above argument, we conclude that A^r satisfies $AA^rA = A$. Further, by the structure of P_1', we see that $P_1'L_1$ has as its non-zero rows exactly the rows of L_1, which are linearly independent. Hence rk $P_1'L_1$ = rk L_1 = s = rk A. Thus, by Theorem IV.3., $P_1'L_1$ is an r-inverse of A. Q.E.D.

The following corollary is also due to Frame.

Corollary V.1.

Let the equation $Ax = y$ be consistent where A is an n x m matrix of rk s. Then the most general solution is given by

$$x = P_1'L_1y + (P_2' - P_1'V)z$$

where z is an arbitrary (m-s)-vector.

Proof

All we need prove is that every vector $(I-A^gA)z_1$ can be written as $(P_2' - P_1'V)z$. Since $A(I-A^gA) = 0$, we have that $C(I-A^gA) = 0$.

Let us then determine the form of those matrices S such that $CS = 0$. We have that $C = [I_r|V]P$. Hence the $m\times(m-s)$ matrix $A_0 =$

$P' \begin{bmatrix} -V \\ I_{m-s} \end{bmatrix}$ is such that $CA_0 = 0$. We then have that the columns of A_0 must belong to $N(C)$. But $\text{rk } A_0 = m-s = \dim N(C)$ since $\dim N(C) + \text{rk } C = m$. Thus, column space $A_0 = N(C)$. Now, since column space $(I - A^gA) \subset N(C) =$ column space A_0, there is an $(m-s)$ vector z such that $(I-A^gA)z_1 = A_0z = (P_2' - P_1'V)z$. Q.E.D.

To help clarify the ideas of Frame's method, we present the following example.

Consider the system of equations

$$
\begin{aligned}
2x_1 + 4x_2 + x_3 + 4x_4 + 4x_5 &= -1 \\
x_1 + 2x_2 \qquad\quad + 2x_4 + x_5 &= 0 \\
x_3 \qquad\quad + x_5 &= 1 \\
2x_1 + 4x_2 + 2x_3 + 4x_4 + 5x_5 &= 0
\end{aligned}
$$

which we write as $Ax = y$.

$$
\begin{bmatrix} 2 & 4 & 1 & 4 & 4 \\ 1 & 2 & 0 & 2 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 2 & 4 & 2 & 4 & 5 \end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}
=
\begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}
$$

Writing $[A|y|I_4]$ gives

$$
\left[\begin{array}{ccccc|c|cccc}
2 & 4 & 1 & 4 & 4 & -1 & 1 & 0 & 0 & 0 \\
1 & 2 & 0 & 2 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
2 & 4 & 2 & 4 & 5 & 0 & 0 & 0 & 0 & 1
\end{array}\right]
$$

Row reduction yields

$$\left[\begin{array}{ccccc|c|cccc} 1 & 2 & 0 & 2 & 0 & -2 & -1 & 3 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 3 & -1 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 & -2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 1 \end{array}\right]$$

Hence we see that the system is consistent. Now,

$$C = \begin{bmatrix} 1 & 2 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}; \quad P' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

so

$$P = \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right] = \left[\begin{array}{c} P_1 \\ \hline P_2 \end{array}\right]$$

$$L_1 = \begin{bmatrix} -1 & 3 & 1 & 0 \\ -1 & 2 & 2 & 0 \\ 1 & -2 & -1 & 0 \end{bmatrix} \quad ; \quad V = \begin{bmatrix} 2 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Hence

$$A^r = P_1{}'L_1 = \begin{bmatrix} -1 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & -2 & -1 & 0 \end{bmatrix}$$

and

$$P_2{}' - P_1{}'V = \begin{bmatrix} -2 & -2 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Hence the most general solution to $Ax = y$ is

$$x = \begin{bmatrix} -1 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & -2 & -1 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -2 & -2 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$$

where $z = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$ is arbitrary.

We should comment here that although we may completely analyze the system of linear equations $Ax = y$ through the use of g.i.'s, this is often not the best way of doing so. There is a well-established method for determining whether or not the system $Ax = y$ is consistent. This method (Ref. 17) consists of adjoining the column vector y to the matrix A to obtain $[A|y]$, and row-reducing this "augmented" matrix. The system $Ax = y$ will be consistent if and only if $\text{rk}\,[A|y] = \text{rk}\,A$, and when it is consistent, we may find a general solution as is shown in (Ref. 17). The case where the g.i.'s are of practical use is when the system $Ax = y$ is inconsistent. We then work with $A'Ax = A'y$, and we may use Rao's method for finding the g.i. of $A'A$ or a g.i. of $(A'A)^2$. We may then easily compute A^+.

NOTATION

$\Longrightarrow$	implies
$\Longleftrightarrow$	is equivalent to
ϵ	belongs to, or is a member of
$\subset$	is a subset of
$\cup$	set union
$\cap$	set intersection
$\{X_i\}_1^n$	the collection $\{X_1, .., X_n\}$.
g.i.	generalized inverse
A, B, X, Y	capital letters to denote matrices
I_n	the nxn identity matrix
x, y, z	small letters at end of alphabet denote vectors.
α, a, b, c	small letters at beginning of alphabet or small greek letters denote real numbers.
proj (y; **W**)	the orthogonal projection of the vector y on the subspace **W**.
(x, y)	the inner product of the vector x with the vector y.

REFERENCES

1. Albert, A. and Sittler, R. W., 1963, "Projection and Pseudo-Inverses Made Computationally Easy". The Arcon Corporation, Lexingon, Mass.

2. Albert, A., 1964, "An Introduction and Beginner's Guide of Matrix Pseudo-Inverses". Arcon Research Project No. RR003-0401, Lexington, Mass.

3. Bellman, R. D., 1960, "Introduction to Matrix Analysis", McGraw-Hill Book Co., New York, N. Y.

4. Ben-Israel, A. and Charnes, A., "Contributions to the Theory of Generalized Inverses", SIAM Journal 11, (3), 1963, pp. 667-699.

5. Cline, R. E., "Note on the Generalized Inverse of the Product of Matrices", SIAM Revue Journal, 6, (1), 1964, pp. 57-58.

6. Cline, R. E. "Representations for the Generlized Inverse of a Partitioned Matrix", SIAM Journal, 1964, 12, (3), pp. 588-600.

7. Decelle, H. P., A Special Form of a Generalized Inverse of an Arbitrary Complex Matrix, NASA TND-2748, 1965.

8. Decelle, H. P., An Application of Generalized Matrix Inversion to Sequential Least Squares Parameter Estimation, NASA TND-2830, 1965.

9. Desoer, C. A. and Whalen, B. H., "A Note on Pseudo-Inverses", SIAM Journal, 1963, 11, (2), pp. 442-447.

10. Foulis, D. J. "Relative Inverses In Baer*-semi-groups", Michigan Math. Journal, 1963, 10, (1), pp. 65-84.

11. Frame, J. S., "Matrix Operations and Generalized Inverses, IEEE Spectrum, March 1964.

12. Friedricks, K. O., "Pseudo Inverses; Completely Continuous Transformations", N. Y. U. Institute of Math. Sciences, 1949.

13. Greville, T. N. E., "The Pseudo-Inverse of a Rectangular or Singular Matrix and Its Application to the Solution of Systems of Linear Equations", SIAM Review, 1, (1), 1959, pp. 38-43.

14. Greville, T. N. E., "Some Applications of the Pseudo-Inverse of a Matrix", SIAM Review, 1960, 2, (1), pp. 15-22.

15. Halmos, P. R., Introduction to Hilbert Space and the Theory of Spectral Multiplicity, Chelsea, New York, 1957.

16. Halmos, P. R., Finite Dimensional Vector Spaces, D. Van Nostranc Co., Inc. Princeton, N. J., 1958.

17. Hoffman, K. and Kunze, R., Linear Algebra, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1961.

18. Moore, E. H., Abstract, Bull. of Amer. Math. Soc., 1920, 26, pp. 394-395.

19. Moore, E. H. "General Analysis", Memoirs of the Amer. Philosophical Soc., 1935, Vol. I.

20. Penrose, R. "A Generalized Inverse for Matrices", Proc. of the Cambridge Philosophical Society, 1955, 51, pp. 406-413.

21. Penrose, R., "On Best Approximate Solutions of Linear Matrix Equations", Proc. of the Cambridge Philosophical Society, 1955, 52, pp. 17-19.

22. Price, C. M. "The Matrix Pseudo-Inverse and Minimal Variance Estimates", SIAM Review, 6, (2), 1964, pp. 115-120.

23. Rao, C. R. "A Note on a Generalized Inverse of a Matrix with Applications to Problems in Mathematical Statistics", Journal of the Royal Stat. Soc., 1962, Series B, 24, (1), pp. 152-158.

24. Rao, M. M. and Chipman, J. S., "Projections, Generalized Inverses, and Quadratic Forms", Journal of Math. Analysis and Applications, 1964, 9, (1), pp. 1-11.

25. Rohde, C. A., "Contributions to the Theory, Computation, and Application of Generlized Inverses", Institute of Statistics Mimeograph Series No. 392, 1964, North Carolina State College, Raleigh, N. C.